

A PDE-CONSTRAINED OPTIMIZATION APPROACH TO
UNCERTAINTY QUANTIFICATION IN INVERSE PROBLEMS
WITH APPLICATIONS TO INVERSE SCATTERING

AFOSR GRANT FA9550-07-1-0480

Omar Ghattas (PI)

Institute for Computational Engineering and Sciences, The University of Texas at Austin

George Biros (Co-PI)

College of Computing, Georgia Institute of Technology

Abstract

This project addresses the statistical inverse problem of reconstruction of an uncertain shape of a scatterer or properties of a medium from noisy observations of scattered wavefields. The Bayesian solution of this inverse problem yields a *posterior pdf*, requiring the solution of the forward wave equation to evaluate the density for any point in parameter space. The standard approach is to sample this pdf via an MCMC method and then compute statistics of the samples. However, standard MCMC methods view the underlying parameter-to-observable map as a black box, and thus do not exploit its structure, hence becoming prohibitive for high dimensional parameter spaces and expensive simulations.

We have developed a Langevin-accelerated MCMC method for sampling high-dimensional PDE-based probability densities. The method builds on previous work in Langevin dynamics, which uses gradient information to guide the sampling in useful directions, improving convergence rates. We have extended the Langevin idea to exploit local Hessian information, leading to a stochastic version of Newton's method. We have also begun to analyze the spectral structure of the Hessian for inverse scattering problems. Applications to model inverse medium scattering problems indicate several orders of magnitude improvement over a reference black-box MCMC method.

Background

The overall goal of this project is to create systematic, rigorous, and scalable algorithms for quantifying uncertainties in inverse wave scattering problems. These uncertainties reflect our incomplete knowledge of the medium in which the waves propagate (inverse medium scattering problem) or the shape of a scatterer (inverse shape scattering problem). The problem of inferring an uncertain medium or shape from observations of scattered wavefields is fundamentally a *statistical inverse problem*. Our lack of knowledge results from noisy measurements, sparse observers, uncertain forward models, and uncertain prior model parameter information. Uncertainty in the reconstructed model parameters is a fundamental feature of ill-posed inverse problems.

The deterministic approach to the inverse scattering problem, which amounts to minimizing a regularized data misfit function, is incapable of accounting for uncertainties in the

20120918187

solution of the inverse problem. *Bayesian inference* provides a systematic framework for incorporating uncertainties in observations, forward models, and prior knowledge to quantify uncertainties in the model parameters. Suppose the relationship between output observables y (such as waveforms at sensor locations) and uncertain model parameters p (such as those describing a wave speed of a heterogeneous medium or shape of a scatterer) is denoted by $y = f(p, e)$, where e represents noise due to measurement and/or modeling errors. In other words, given the model parameters p and noise e , the function $f(p, e)$ solves the forward (acoustic, elastic, or electromagnetic wave propagation) problem to yield y , the predicted outputs at the measurement locations (and time instants). Suppose also that we have the prior probability density $\pi_{\text{prior}}(p)$, which encodes the confidence we have in prior information on the unknown model parameters (i.e. independent of present observations), and the likelihood function $\pi(y_{\text{obs}}|p)$, which describes the conditional probability that the model parameters p gave rise to the actual measurements y_{obs} . Then Bayes' theorem of inverse problems expresses the posterior probability density of the model parameters, π_{post} , given the data y_{obs} , as the conditional probability

$$\pi_{\text{post}}(p) \stackrel{\text{def}}{=} \pi(p|y_{\text{obs}}) \propto \pi_{\text{prior}}(p) \pi(y_{\text{obs}}|p). \quad (1)$$

Expression (1) provides the statistical solution of the inverse problem as a probability density for the model parameters p .

While it is easy to write down the expression (1) for the posterior probability density, making use of this expression poses a challenge, because the posterior probability density is a surface in high dimensions (equal to the number of model parameters p), and because the solution of the forward problem (i.e., computing $f(p)$ given p) is required to evaluate the probability of any point in parameter space. Straightforward grid-based sampling is out of the question for anything other than a few parameters and cheap forward simulations. Special sampling techniques, such as Markov chain Monte Carlo (MCMC) methods, have been developed to generate sample ensembles that typically require many fewer points than grid-based sampling. Still, when the model parameters represent a (suitably-discretized) field (such as scatterer shape or medium wave speed), and when the forward PDE requires hours to solve on a parallel computer (such as mid-to-high frequency wave propagation), the MCMC framework becomes completely intractable.

The central difficulty in scaling up conventional MCMC for large-scale forward simulations and high-dimensional parameter spaces is that this is a purely *black-box* approach, i.e. it does not exploit the structure of the parameter-to-observable map $f(p)$. Twenty years of advances in algorithms for *deterministic* large-scale PDE-constrained optimization have taught us that making maximal use of derivative information can greatly speed up the search process for extremum points. The goal of this project is to overcome the intractability of conventional methods for statistical inverse scattering problems by developing scalable algorithms that exploit the structure of inverse wave propagation operators. Our work this year has focused on developing preconditioned Langevin methods that exploit this structure to greatly speed up sampling. This has involved the analysis and construction of fast algorithms for approximating the inverse Hessian for inverse wave scattering problems.

Status/Progress: Fast Hessian-preconditioned Langevin samplers (UT Austin)

This past year we have been developing fast sampling methods that build on—and significantly extend—ideas from Langevin dynamics, which use gradient information to accelerate sampling of a target density. The Langevin equation is a stochastic differential equation,

$$d\mathbf{P}_t = \mathbf{A} \nabla \log \pi_{\text{post}} dt + \sqrt{2} \mathbf{A}^{1/2} d\mathbf{W}_t, \quad (2)$$

with $\pi_{\text{post}}(p)$ as an invariant density. Here, \mathbf{W}_t is the i.i.d. vector of standard Brownian motions. Preconditioning by a symmetric positive definite operator \mathbf{A} preserves the invariance of the density. In practice, we discretize in time with timestep Δt , yielding (e.g. for explicit Euler) the update

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{A} \nabla \log \pi_{\text{post}} \Delta t + \sqrt{2\Delta t} \mathbf{A}^{1/2} \mathcal{N}(0, \mathbf{I}) \quad (3)$$

where $\mathcal{N}(0, \mathbf{I})$ is the i.i.d. standard normal density. Discretization in time can add bias, so we use the Langevin steps as proposals for the Metropolis-Hastings algorithm. The form (3) shows immediately the connection with deterministic optimization methods: the gradient term $\nabla \log \pi_{\text{post}}$ is a steepest ascent direction for the posterior density. In its absence (and in the absence of preconditioning, i.e. $\mathbf{A} = \mathbf{I}$) we recover a Gaussian random walk from the last term in (2). The addition of the gradient term drives the sampling in (the locally steepest) direction of higher probability. However, steepest descent is a poor choice for search directions in large-scale optimization (particularly for anisotropic pdfs), and we seek to improve on it.

Taking the preconditioner \mathbf{A} as the inverse of the Hessian of $\log \pi_{\text{post}}$, we obtain the stochastic equivalent of Newton's method. In the common case of Gaussian additive noise and prior, the (negative) log of the posterior density is simply the “regularized” misfit function (the sum of the data misfit and prior/regularization term) that deterministic inverse methods seek to minimize. Thus, similar to Newton's locally-quadratic approximation of the objective, the Hessian-preconditioned Langevin step makes a locally-Gaussian approximation of π_{post} . This endows the sampling process with information on the curvature of the posterior density surface, which is crucial in high dimensions. We expect this to result in a need for substantially fewer sampling points relative to a black-box MCMC method, just as deterministic Newton requires substantially fewer iterations to find the optimum compared to a derivative-free optimization method.

Moreover, it can be shown that in the limiting case when the posterior density π_{post} is in fact Gaussian (e.g. when the inverse problem is linear and the noise is additive and Gaussian), this so-called *stochastic Newton* method not only samples the target density at long times, but accurately samples from π_{post} at every time step. This means that the Metropolis-Hastings criterion will accept *all* of the proposed sample points, and that a minimum number of points are needed to accurately sample from the given distribution. For densities that are not Gaussian, stochastic Newton will still provide a substantial speedup over a conventional random walk, since a Gaussian approximation (based on a local quadratic approximation of $\log \pi_{\text{post}}$, or equivalently a linearized approximation of the inverse problem) will generally prove to yield more useful information on the behavior of π_{post} than a standard normal density approximation (or other heuristic) will.

We have developed a preliminary implementation of the stochastic Newton method, and applied it to solve nonlinear inverse medium and shape scattering problems in one and two dimensions. For example, for a 65-parameter 1D inverse medium problem, Figure 1 indicates just $\mathcal{O}(10^2)$ samples are necessary to adequately sample the (non-Gaussian) posterior density, while a reference (non-derivative) MCMC method (Delayed Rejection Adaptive Metropolis (DRAM)) is nowhere near converged after even $\mathcal{O}(10^5)$ samples. The performance of unpreconditioned Langevin MCMC is similar to that of DRAM, indicating the crucial role of the Newton direction vs. steepest descent. Moreover, because the (inverse) Hessian captures the (local) covariance structure of the posterior density, this orders-of-magnitude speedup is expected to become even larger as the parameter dimension increases.

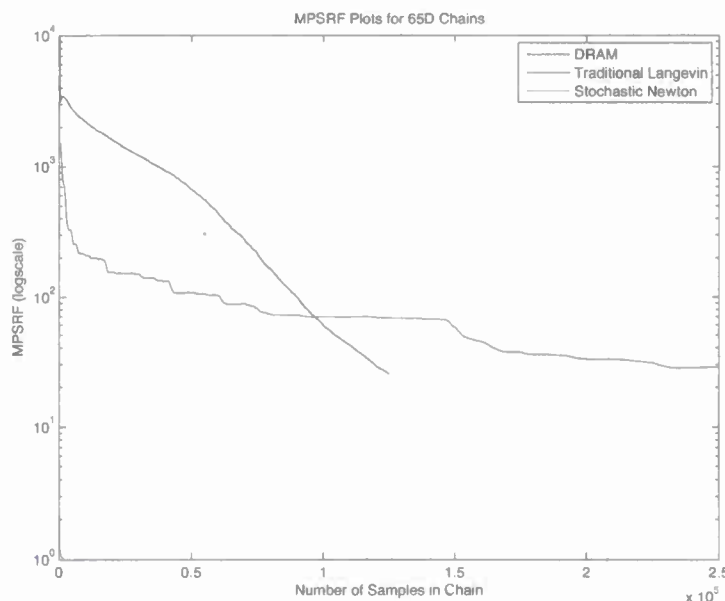


Figure 1: Comparison of number of points taken for sampling posterior density for a 65-dimensional inverse scattering problem to identify the distribution of elastic moduli for a layered medium, from reflected waves. DRAM (black), unpreconditioned Langevin (blue), and Stochastic Newton (red) sampling methods are compared. Convergence indicator is multivariate potential scale reduction factor (MPSRF) for which a value of unity indicates convergence. Stochastic Newton requires *three orders of magnitude* fewer sampling points.

Status/Progress: Fast approximations of Hessian operators (Georgia Tech)

This past year we have also been working on fast approximations for Hessian operators, which are critical component of fast sampling algorithms for statistical inverse problems. Hessian operators are mainly determined by three characteristics of the inverse problem: the parameters, the underlying PDE, and the observation operator. Previously we have analyzed Hessians for definite elliptic and parabolic equations, for different observation operators for the inverse source and inverse medium problems (see publications). Recently we have started working with indefinite elliptic operators, namely the Helmholtz operator for inverse medium problems. We consider the problem of the inversion of the scalar wave

equation in a lossy medium. In particular, we focus on broadband multi-point illumination problems for low frequency regimes.

More specifically, we consider

$$-\nabla^2 u(\mathbf{r}, t) + \sigma(\mathbf{r}) \frac{\partial u(\mathbf{r}, t)}{\partial t} + \frac{1}{c^2(\mathbf{r})} \frac{\partial^2 u(\mathbf{r}, t)}{\partial t^2} = S(\mathbf{r}, t) \quad (4)$$

where u denotes the state, c the speed of sound, σ viscous dissipation, and S the source term. If all variables are time-harmonic with a fixed angular frequency ω , in the frequency domain Equation (4) becomes

$$-\nabla^2 u(\mathbf{r}) + \left[-i\omega\sigma(\mathbf{r}) - \frac{\omega^2}{c^2(\mathbf{r})} \right] u(\mathbf{r}) = S(\mathbf{r}, \omega). \quad (5)$$

We wish to design a fast method for the recovery of either $c(\mathbf{r})$ or $\sigma(\mathbf{r})$ from the values of the scattering data measured on N_d detectors. To generate the set of scattered measurements, a small number of point sources (N_s) will be considered. The positions of the sources will be denoted \mathbf{r}_s^i ($1 \leq i \leq N_s$). For each source, we consider N_f frequencies ω .

We use an integral equation formulation for the forward problem and a Born approximation for the inverse problem, leading to an ill-posed linear least-squares problem. If N_f is the number of excitation frequencies, N_s the number of different locations for the point illuminations, N_d the number of detectors, and N the parametrization for the scatterer, a direct SVD will have $O((N_s N_d N_f)^2 N)$ computational cost. We have developed a fast SVD method that, by introducing a controllable error in the factorization, brings this cost down to $O(N_f N)$ thus, providing orders of magnitude speed-up over the generic factorization algorithm. The fast SVD approach consists of the following components: a plane wave expansion of the forward operator; a fast randomized SVD for the single-source-many-frequencies problem; and a multi-step combination of the single source SVDs to construct the overall factorization. We are currently verifying and validating our method and are examining ways to extend it to higher frequency regimes.

Acknowledgment/Disclaimer

This work was sponsored (in part) by the Air Force Office of Scientific Research, USAF, under grant number FA9550-07-1-0480. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

Personnel Supported During Duration of Grant

- Omar Ghattas, Professor, UT-Austin
- George Biros, Professor, Georgia Tech
- Tan Bui-Thanh, Postdoc, UT-Austin
- Stephanie Chaillat, Postdoc, Georgia Tech
- James Martin, Graduate Student, UT-Austin

Publications

1. S.S. Adavani and G. Biros, *Fast Algorithms for Source Identification Problems with Parabolic PDE Constraints*, submitted, (26pp) 2009
2. S.S. Adavani and G. Biros, *Fast Algorithms for Source Identification Problems with Elliptic PDE Constraints*, **SIAM Journal on Imaging Sciences**, (in press), (24pp) 2009
3. S.S. Adavani and G. Biros, *Multigrid algorithms for inverse problems with linear parabolic PDE constraints*, **SIAM Journal on Scientific Computing**, 31(1), 369–397, 2008.
4. A. Askan, V. Akcelik, J. Bielak, and O. Ghattas, *Parametric analysis of a nonlinear least squares optimization-based anelastic full waveform inversion method*, submitted.
5. G. Biros and G. Doğan, *A multilevel algorithm for inverse problems with elliptic PDE constraints*, **Inverse Problems**, 24(3), (19pp), 2008.
6. C. Burstedde and O. Ghattas, *Algorithmic Strategies for Full Waveform Inversion: 1D Experiments*, **Geophysics**, to appear.
7. C. Burstedde, O. Ghattas, J. Martin, and L.C. Wilcox, *Uncertainty Quantification in Inverse Problems with Stochastic Newton MCMC*, in preparation, 2009.
8. S. Chaillat and G. Biros, *A fast algorithm for multi-frequency and multipoint inverse scattering problems*, in preparation, 2009.
9. I. Epanomeritakis, V. Akcelik, O. Ghattas, and J. Bielak, *A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion*, **Inverse Problems**, 24:034015 (26pp), 2008.

Honors and Awards

- The paper *A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion* (paper 9 above) was included in the 2008 Highlights of the journal **Inverse Problems**.
- The paper *C. Burstedde, O. Ghattas, M. Gurnis, E. Tan, T. Tu, G. Stadler, L. Wilcox, S. Zhong, Scalable adaptive mantle convection simulation on petascale supercomputers*, *Proceedings of SC08* was a finalist for the 2008 Gordon Bell Prize.
- PI Ghattas gave the keynote talk at the *10th LCI International Conference on High-Performance Clustered Computing* in Boulder in March 2009. Co-PI Biros gave an invited plenary talk at *Applied Inverse Problems* in Vienna in July 2009.

AFRL Point of Contact: none

Transitions: none

New Discoveries: none

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY)
12-03-20102. REPORT TYPE
Final Technical3. DATES COVERED (From - To)
01 June 2007 - 30 Nov 20094. TITLE AND SUBTITLE
A PDE-CONSTRAINED OPTIMIZATION APPROACH TO UNCERTAINTY IN
INVERSE PROBLEMS WITH APPLICATIONS TO INVERSE SCATTERING

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-07-1-0481

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
George Biros

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
School of Computational Science and Engineering
Georgia Institute of Technology8. PERFORMING ORGANIZATION
REPORT NUMBER9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/RSL
875 N Randolph St, Suite 325
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT
NUMBER(S)

AFRL-OSR-VA-TR-2012-0021

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Our goal is the design of fast parallel algorithms statistical inference for scalar and wave propagation problems. We have looked at source inversion and inverse medium problem problems. We use a Bayesian approach in which the regularization appears as prior information and the data mismatch appears as a likelihood information, given known noise probability density functions.

A key component of all of our algorithms is the approximation of the Hessian operator. Key components of our work are rank-revealing factorizations, fast extraction of the diagonal of the inverse, adaptivity, and integration of all of these components within a particle filter methodology.

In addition, our implementations are being designed to scale on manycore and heterogeneous parallel architectures.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:

a. REPORT b. ABSTRACT c. THIS PAGE

UU

UU

UU

17. LIMITATION OF
ABSTRACT

UU

18. NUMBER
OF
PAGES

4

19a. NAME OF RESPONSIBLE PERSON

Fariba Fahroo

19b. TELEPHONE NUMBER (Include area code)

7036968429